ESTIMATION OF TARGET RANGE AND ORIENTATION
FROM VIDEO IMAGES

August 2001

Anthony G. Galaitsis and Patrick J. Coskren
BBN Technologies
10 Moulton Street
Cambridge, MA 02138

ABSTRACT

Under National Ground Intelligence Center (NGIC) sponsorship, BBN Technologies (BBN) has initiated the development of a model to estimate the acoustic source strength of a target at different azimuth, elevation, and viewing-axis angles from a videotaped record. The model comprises two distinct parts: a) Estimation of the target range and orientation from video images and known geometric information; b) Estimation of the acoustic source strength from recorded signatures corresponding to the video images. The first part of this work has been completed and is the subject of this paper.

The Target Range and Orientation Estimation Model ("Video Range Finder") is based on an interactive procedure that enables the user to associate identifiable features of the video image with distinct, pre-selected physical components of the target. These points define a 3-dimensional polyhedron in the coordinate frame of the target, which is projected onto the 2-dimensional polygon image at the focal plane of video camera, along the viewing axis. The azimuth, elevation, and viewing-axis rotation angles are obtained through an LMS solution of the associated projection equations, which are non-linear with respect to these angles. The calculated angles are subsequently used in the estimation of the target range. The paper discusses the solution method and demonstrates the use of the model using both simulated and real objects.

Report Documentation Page

Report Date 01AUG2001	Report Type N/A	Dates Covered (from... to) -

Title and Subtitle Estimation of Target Range and Orientation from Video Images	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) BBN Technologies 10 Moulton Street Cambridge, MA 02138	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) Department of the Army, CECOM RDEC Night Vision & Electronic Sensors Directorate AMSEL-RD-NV-D 10221 Burbeck Road Ft. Belvoir, VA 22060-5806	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement Approved for public release, distribution unlimited

Supplementary Notes Papers from 2001 Meeting of the MSS Specialty Group on Battlefield Acoustic and Seismic Sensing, Magnetic and Electric Field Sensors, Volume 1: Special Session held 23 Oct 2001. See also ADM001434 for whole conference on cd-rom., The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified

Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages 11

1. INTRODUCTION

One of NGIC's missions is to gather and organize acoustic signatures of foreign vehicles, and make them available to the defense community. Real vehicles are characterized by complex geometry, and they, generally, have multiple acoustic sources onboard that result in significant directivity patterns. Therefore, it is important that source range and orientation information be developed and stored along with the signatures in the database, and that it be accounted for during retrieval and utilization of the data.

Currently, the vehicle range and orientation information is estimated through manual scaling of video vehicle images acquired concurrently with acoustic signatures. This manual procedure has been cumbersome, therefore, there has been a need for an enhanced process that will improve the data analysis accuracy and speed up the data availability to the defense community.

The objective of this work was to develop an interactive data analysis tool that minimizes the required manual user intervention and generates the desired target range and orientation information accurately and efficiently.

2. TECHNICAL APPROACH

The target range and orientation are obtained through a parameter estimation method that solves a set of non-linear equations relating these quantities to specific features of the target and of acquired video images. This section discusses the highlights of this approach.

The Problem: The key technical issues involved in the estimation of the range and orientation of an object from a video image, and from previously known geometric information about the object are illustrated in Figure 1.

Figure 1a shows how an object of true length L and located at a distance D from a camera is converted to an image of height h on the focal plane of the camera and, subsequently, into a digital image of p pixels on a computer monitor. The process involves three distinct mappings, namely:

a) A projection of the object onto the PP-plane, which converts the 3-dimensional object of true length L into a 2-dimensional object of length H,

b) An Image Size Reduction through the camera lens that converts the projected length H into the reduced length h at the focal plane of the camera, and

c) An analog-to-digital mapping that converts the "film" image of length h into the digital image of p pixels

The composite mapping is also illustrated qualitatively in Figure 1b, where a 3-dimensional vector length **DV** = [DY, DY, DZ], expressed in meters, feet, or other length units**,** is converted to a 2-dimensional digital image **dv** = [dx, dy]**,** expressed in pixels.

Projection Transformation: The projection of an object onto the PP-plane is illustrated in Figure 2 by the two coordinate systems defined by the X-Y-Z and x-y-z axes. In Figure 2a the z-y plane is rotated by an azimuth angle ϕ about the Z-axis. In Figure 2b the z'-axis is elevated by an angle θ with respect to the X-Y plane (Figure 2b). Assuming that the camera-viewing
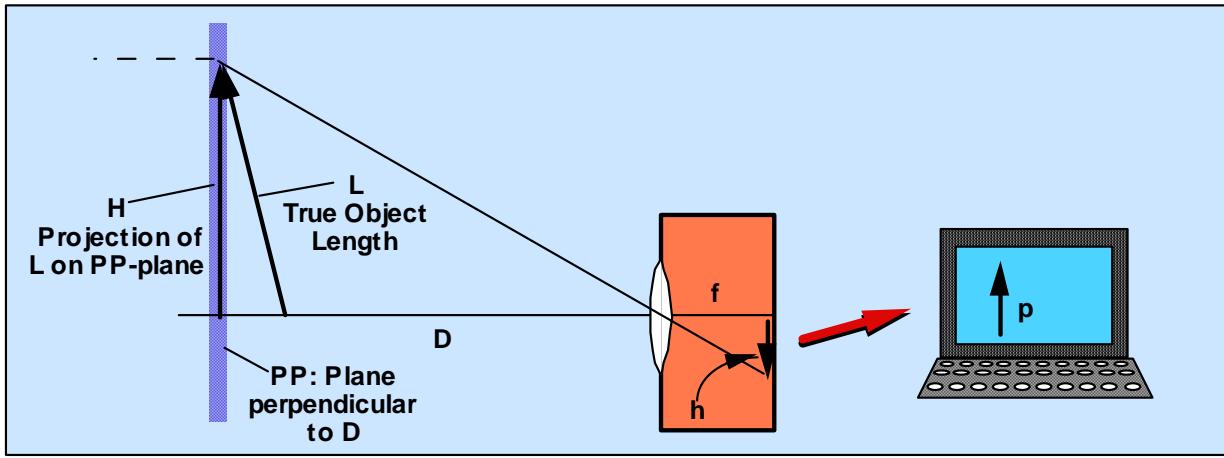
(a)
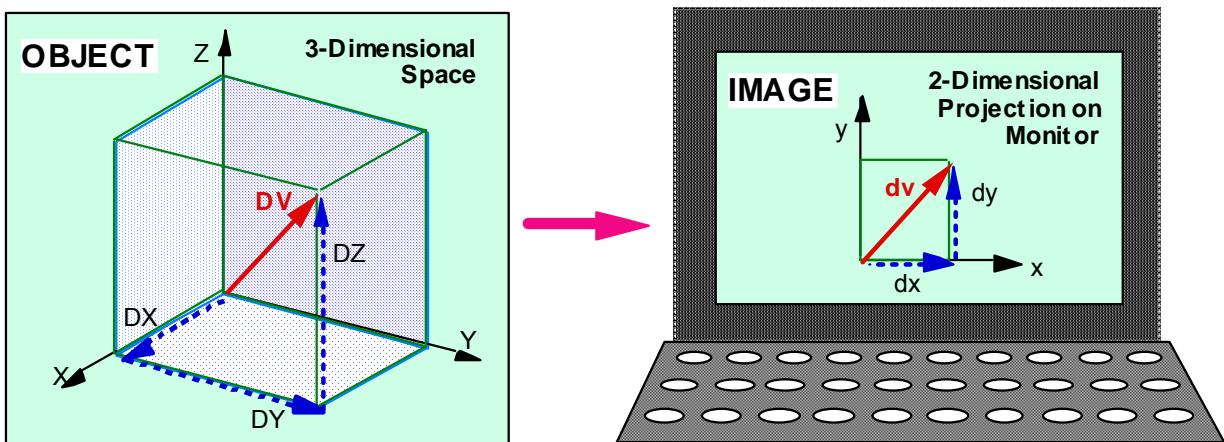
(b)

FIGURE 1. Transformation of a 3-Dimensional Object into a 2-Dimensional Digital Image

direction is along the negative z-axis, the x'-y' plane of Figure 2b corresponds to the PP-plane of Figure 1a. It follows that a 2-dimensional projection $\mathbf{r'_i} = [x'_i, y'_i]^T$ onto the x'-y' plane of a 3-dimensional vector length $\mathbf{R_i} = [X_i, Y_i, Z_i]^T$ in the X-Y-Z space is given by:

$$\mathbf{r'_i} = \mathbf{H}(\theta,\phi)\mathbf{R}_i \tag{1}$$

where $\mathbf{R_i} = [X_i, Y_i, Z_i]^T = \text{transpose}[X_i, Y_i, Z_i]$ is a column vector, and the transformation matrix $\mathbf{H}(\theta,\phi)$ is defined by:

$$\mathbf{H}(\theta,\phi) = \begin{bmatrix} -\sin(\phi) & \cos(\phi) & 0 \\ -\cos(\phi)\sin(\theta) & -\sin(\phi)\sin(\theta) & \cos(\theta) \end{bmatrix} \tag{2}$$

Typically, the camera is also rotated through an angle ω about the viewing axis, as indicated in Figure 2c, which is accounted for by the rotation matrix

$$\mathbf{G}(\omega) = \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix} \tag{3}$$

Therefore, the coordinates, $\mathbf{r''_i} = [x''_i, y''_i]^T$, of a projection onto the PP-plane (or x"-y" plane) of the vector $\mathbf{R_i}$ is given by:

$$\mathbf{r''_i} = \mathbf{G}(\omega)\,\mathbf{H}(\theta,\phi)\,\mathbf{R}_i \tag{4}$$

Image Size Reduction: This mapping performed by the camera lens performs an optical image reduction given by the ratio q = h/H (Figure 1a) or, equivalently, by the ratio

q = d/D

where D is the range to the object and d is the distance between the camera lens and its focal plane (Figure 1a).

Analog-to-Digital Mapping: At the focal plane of the digital camera, the reduced 2-dimensional projection of the object is converted into a digital image and, subsequently, represented by an array of pixels. This transformation is expressed by the parameter

B: Pixels per unit length (5)

Generally, in addition to the rotation and magnification effects covered in the above transformations, there is also a translation between the origins of any two coordinate systems. This is accounted for by introducing a fixed offset term. The final expression relating the size of the 3-dimensional object and its 2-dimensional digital image is:

$$\boldsymbol{\rho}_i = \boldsymbol{\rho}_o + (Bd/D)\,\mathbf{G}(\omega)\,\mathbf{H}(\theta,\phi)\,\mathbf{R}_i \tag{6}$$

where $\boldsymbol{\rho}_i = [\rho_{ix}, \rho_{iy}]$ are the image coordinates in pixels, and the fixed offset $\boldsymbol{\rho}_0$ accounts for the different origins of the two systems.
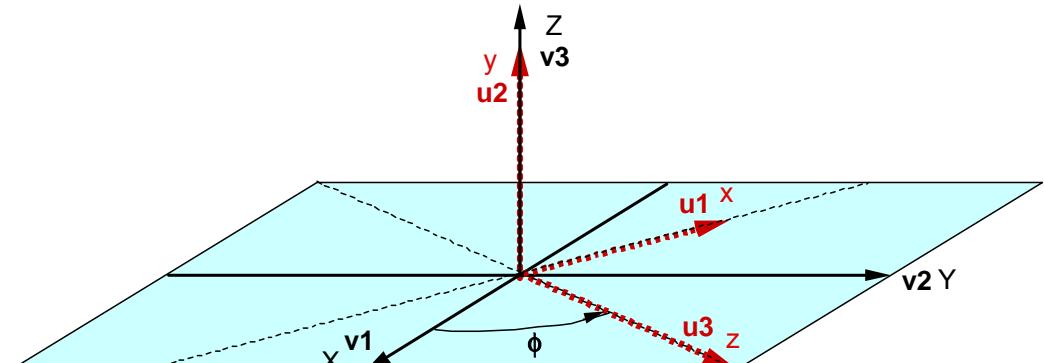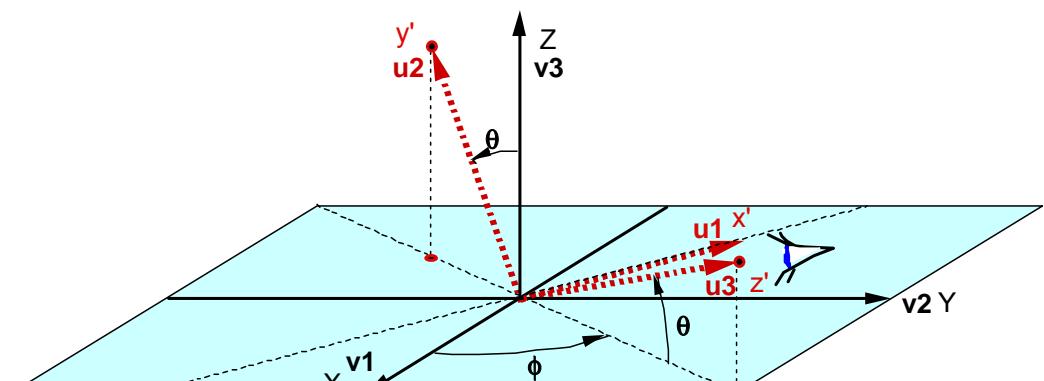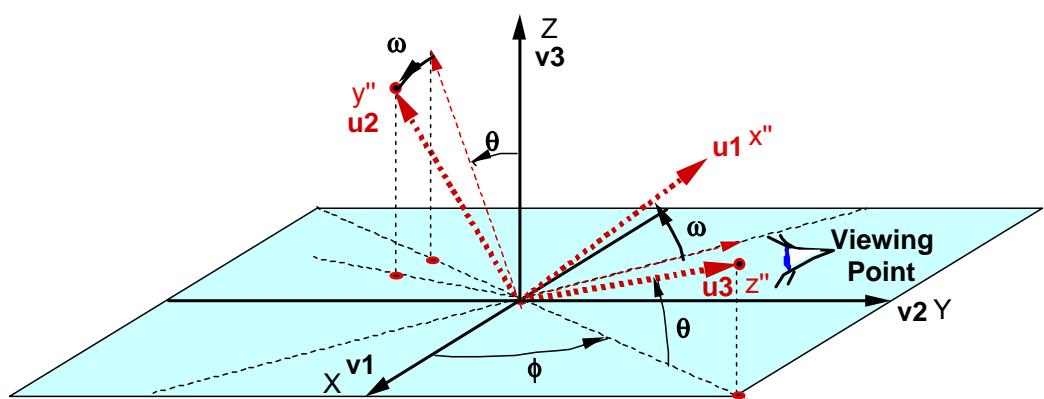
FIGURE 2. **Rotational Transformation Angles: a) Azimuth ϕ, b) Elevation θ, and c) Viewing Axis Angle ω.**

It should be noted that the above relationships neglect the effects of perspective. However, as long as the range is greater than five object diameters, the effect of this approximation is negligible. This validity of this assumption was supported by measurements, as discussed in the next section.

Equations 1-6 provide closed-form expressions for calculating the image coordinates (or size) ρ_i, from the object dimensions, $\mathbf{R}_i$, range, D, coordinate transformation angles, ϕ, θ, and ω, and camera/monitor parameter Bd. In the present case, however, the objective is to estimate the range, D, and orientation angles, ϕ, θ, and ω, from the known object dimensions, image dimensions and camera monitor parameters. This is accomplished through the numerical inversion process discussed in the remaining of this section.

The Solution: The range D, and the angles, ϕ, θ, and ω are extracted by solving a simultaneous system of equations containing these unknowns. This is performed by applying Equation 6 to several sets of object/image points, thus, generating one vector equation for each point. Since Equation 6 is a non-linear function of ϕ, θ and ω, the solution is obtained numerically.

 The numerical solution is based on an optimization process that minimizes the cost function

$$\xi = \sum (\mathbf{v}_{ij} - \mathbf{u}_{ij})^2 \tag{7}$$

following the negative gradient to the minimum-value point within the associated performance volume. In this expression, u_{ij} and v_{ij} are the <u>measured</u> and <u>theoretical</u> unit vectors, respectively, between the ith and j^{th} points of the 2-dimensional image. These vectors are obtained from

$$\mathbf{u}_{ij} = [u_{xi}-u_{xj}, u_{yi}-u_{yj}] = [\rho_{xim}-\rho_{xjm}, \rho_{yim}-\rho_{yjm}] / |\boldsymbol{\rho}_{im}-\boldsymbol{\rho}_{jm}| \tag{8}$$

$$\mathbf{v}_{ij} = [v_{xi}-v_{xj}, v_{yi}-v_{yj}] = [\rho_{xit}-\rho_{xjt}, \rho_{yit}-\rho_{yjt}] / |\boldsymbol{\rho}_{it}-\boldsymbol{\rho}_{jt}| \tag{9}$$

where ρ_{im} is the <u>measured</u> 2-dimensional pixel vector of the ith point in the digital image, and ρ_{it} is the corresponding <u>theoretical</u> pixel vector of the ith point as calculated from Equation 6.

Equations 7-9 are range-independent, therefore, the first part of the approach estimates a solution for the parameters ϕ, θ, and ω. This is achieved through an iterative method that finds the minimum of the Equation 7 cost function using a Steepest Decent approach. Specifically, the process begins with an arbitrarily selected initial "guess" solution (ϕ_g, θ_g, and ω_g). This solution is used to evaluate the expression of Equation 7 and of the associated gradient. The search continues using new trial solutions selected along the path of the negative gradient (steepest decent along a generalized performance surface) which eventually converges to the desired solutions, ϕ_s, θ_s, and ω_s, that minimize ξ.

The cost function requires a minimum of four distinct points, which define a non-singular polyhedron in the 3-dimensional space of the object, i.e., with not all points located on a single plane. Upon completion of the of ϕ, θ, and ω estimation, the range is obtained by plugging these parameters in Equation 6. The results of calculations based on several test objects are discussed in the next section.

3. MODEL VALIDATION

The Range and Orientation relations of the previous section were used with a Steepest Decent algorithm to estimate the corresponding solutions for ϕ, θ, and ω, numerically. The algorithm was implemented in MATLAB and validated using simulated objects, small-scale vehicle models and, finally, a real size vehicle.

Simulated Objects Validation

The Video Range Finder model was first tested using simulated 3-dimensional objects and their analytically derived 2-dimensional projections. This procedure is illustrated by the MATLAB generated set of plots displayed in Figure 3.
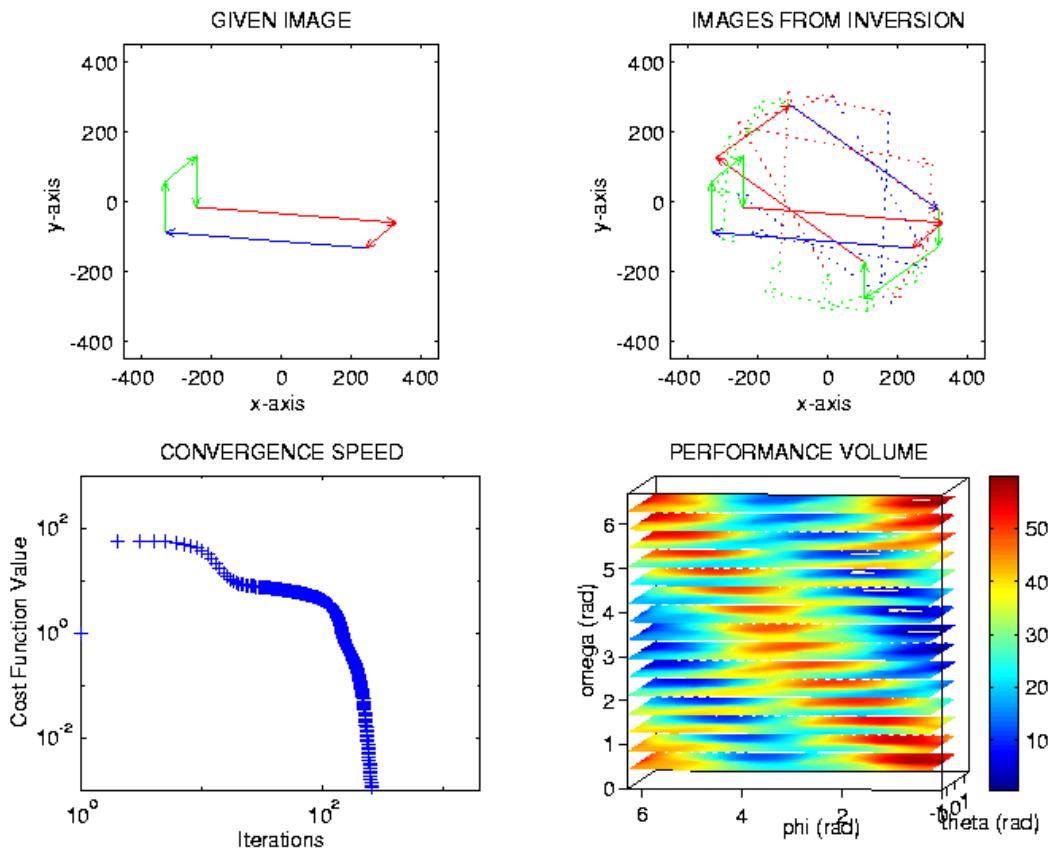
FIGURE 3.   Video Range Finder Validation with a Simulated Object.

Forward Calculations: The first step in this process was to synthesize a hypothetical 3-dimensional polyhedron (a tetrahedron at a minimum), which was defined by the coordinates (expressed in inches) of a set of user-selected points. The 2-dimensional video image (Figure 3a) of this simulated object, i.e., the projections of its apexes and edges onto a plane, was obtained from Equation 6 by assuming that the associated

-- Range D,

-- Projection angles ϕ, θ, and ω,

-- Camera/monitor parameters d, B, and ρ_o

appearing in Equation 6 were all known. The calculated coordinates, ρ_i, of the projected apexes were substituted into Equation 9 to obtain the unit vectors u_{ij} needed in Equation 7.

Inverse Calculations: In the second step, the non-linear equations represented by Equation 6 were solved numerically to <u>recover</u> the original object. Specifically, the range, D_r, and orientation angles ϕ_r, θ_r, and ω_r, of the object were estimated using the known

-- Coordinates, $\rho_i + \varepsilon_i$, on the image,

-- Geometry of the 3-dimensional object, and

-- Camera/monitor parameters d, B, and ρ_o

where ρ_i are the projections obtained from the forward calculations, and ε_i is a user-added noise contribution to simulate measurement errors.

The accuracy of the estimated parameters was checked through direct comparison to the user-selected values and also through reproduction of a projection based on the numerically calculated solution. Figure 3b shows projections corresponding to the guess solution (solid curve polygon), to a few intermediate-step solutions (dotted curves), and to the converged solution (solid curve). A close inspection shows that the given projection (Figure 3a) is visually identical to the projection of the recovered object (Figure 3b).

The evolution of the cost function value during the inverse calculations is shown in Figure 3c. Not surprisingly, the cost function has a relatively high value at the initial trial solutions, but it decreases rapidly after a number of iterations, as it converges towards the actual solution. In the example of Figure 3, the cost function minimum is close to zero because it involves the solution of an ideal case (with assumed $\varepsilon_i = 0$), while in real cases involving some measurement error ($\varepsilon_i \neq 0$), the cost function has a non-zero minimum.

The inversion process requires a minimum of four projected points (not all on the same plane) for an unambiguous solution. In all simulated cases considered, the algorithm converged to a solution that was substantially close to the true solution. This was as expected because the ideal cost function appears to have a single minimum within the $0<\phi<2\pi$, $-\pi/2<\theta<\pi/2$, and $0<\omega<2\pi$. This is illustrated by the Figure 3d visual representation of the sliced cost function volume within the above range of angles, which features a single minimum (dark blue region). The color distribution was different for the different simulated cases considered but, in all cases, a visual inspection of Figure 4d revealed a single minimum.

Small-Scale Model Validation

A second round of validation tests was performed using small-scale models of various vehicles, such as the 12-inch long toy car of Figure 4a.

In this case, measurements were conducted to accurately locate several readily recognizable and compact components, such as headlights, wheel centers, mirrors, windshield corners, muffler exhaust, trunk lid edge center, etc., with respect to a convenient coordinate system embedded in the vehicle [these points are referred to as "designation points"]. After these measurements, the toy car was photographed with a camera from various angles and distances, and the acquired images were transferred to a laptop computer, along with information about the 3-dimensional coordinates of the designation points, and the camera/monitor parameters d, B, and ρ_o.

During execution, the inversion program prompted the user to identify several of the designation points in an imported image. This was performed through mouse-clicks of points on the visible side of the vehicle. The inversion software confirmed the selection of each designation point by drawing a circle around it, as shown in Figure 4b. When the selection was finished, the inversion code computed and reported the range and orientation angles, which were routinely compared to the values read off the protractors (Figure 4a) that were attached to the rotating base of the model.

As in the case of simulated objects, the algorithm converged to a solution (for D, ϕ, θ, and ω) that was consistently close to the true solution. The solution method seemed to be fairly robust with respect to the error associated with an imperfect placement of the cursor on the designation points. Generally, estimation accuracy improved with the number of projection points but began to plateau at about six points.

Full-Scale Vehicle Validation

Finally, the model was validated using photographs of a full-scale vehicle, in this case the Saab-9000 car of Figure 4c.

In this case, several designation points were selected on all sides of the vehicle, and they were marked with white paper stickers, which are clearly visible in Figures 4c. The positions of all designation points were measured with respect to a coordinate frame whose origin was on the ground directly below the right corner of the rear bumper. Subsequently, the Saab was photographed from various positions, with up to ~30 meters distances and several combinations of orientation angles (azimuth, elevation, and viewing axis rotation) to generate a diverse mix of test cases. Also measured were the X-Y-Z coordinates of the camera
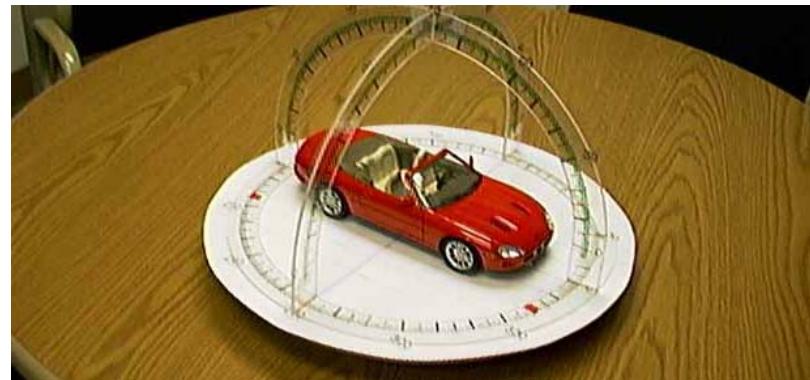
(a)
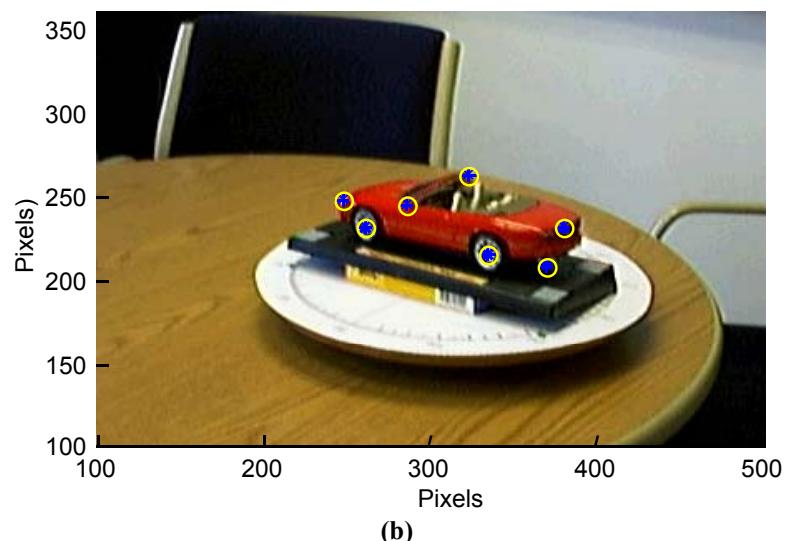
(b)

(c)

FIGURE 4.   Video Range Finder Validation through (a)-(b) Small-Scale Vehicle, and   c) Full-Scale Vehicle.

positions, which were used to obtain the true azimuth and elevation angles (ϕ, θ). The true viewing-axis angle ω, was estimated by the photographer.

The acquired Saab images and the coordinates of the designation points were imported to the laptop computer containing the Video Range Finder software, and the associated code was executed using the full-scale vehicle data. For all cases considered, the Object Range and Orientation calculated by the Video Range Finder code compared very well with the corresponding values that were measured during the outdoors picture taking session.

Errors were typically less the ±5 degrees for the angles and less than ±10% for the range. For ranges 5 times or greater than the typical object length, the errors stemmed primarily from the user's inability to "mouse-click" on the designation points with high precision. At ranges less than 5 object lengths, the errors stemmed primarily from the absence of perspective effects. Errors associated with the convergence of the numerical calculations were judged to be insignificant compared to the above two effects.

4. CONCLUSIONS AND RECOMMENDATIONS

The work performed to date under this NGIC-sponsored program has lead to the development and successful demonstration of a Video Range Finder model. The model has been validated using computer simulations, small-scale model tests, and full-size vehicle tests. Therefore, it is recommended for implementation into NGIC's vehicle signature analysis toolkit.

ACKNOWLEDGEMENT

The authors wish to thank Mr. Robert Grachus (NGIC) for his support and encouragement throughout this project.